Ingeniería Industrial.
Asignatura: Estadística I Año: 3er.
Semestre Ce.
Clase Encuentro # 1. Orientación Tema 1. (4h)

Sumario.
Tema 1 Estadística Descriptiva
1.1 Introducción a los métodos estadísticos. Conceptos Básicos. Ejemplos. Estadística Descriptiva. Observación Discreta y Continua.
1.2 .1 Distribución de frecuencia Univariada. Datos no agrupados. Propiedades de las Frecuencias.
1.2.2. Datos Agrupados. Clases. Límites de clases. Tamaño del Intervalo de Clases. Marcas de Clases.
1.3 Gráficos: Histograma y Polígono de frecuencias (Orientar).
1.4 Medidas Descriptivas de Posición, (Tendencia central): Moda, Mediana y Media (propiedades).
1.5 Medidas Descriptivas de Dispersión: Rango, Varianza, Desviación Típica y Coeficiente de Variación. Propiedades.
1.6 Medidas descriptivas de formas: Asimetría y Curtosis
1.7 Medidas Descriptivas datos cualitativos: Razón, Proporción y Porcentaje.

Objetivos: Que el alumno se capaz de:
- Describir el comportamiento de masas de datos mediante el cálculo e interpretación de medidas descriptivas de tendencia central tales como: moda, mediana y media y de dispersión como rango, varianza, desviación típica y coeficiente de variación.

Método:
- Elaboración conjunta.

- Tiza y pizarra, Libro de texto.
- Forma de enseñanza: Conferencia

Bibliografía:
- Estadística para ingenieros 6ta edición R. Walpole, R. H. Myers, S.L.Myers I (1ra y 2da parte).
- Estadística Dra Caridad Bustillo y otros
- Materiales entregados y elaborados por el profesor

1.1 . Introducción a los métodos estadísticos. Conceptos Básicos. Ejemplos. Estadística Descriptiva. Observación Discreta y Continua

Presentación del profesor, sus objetivos

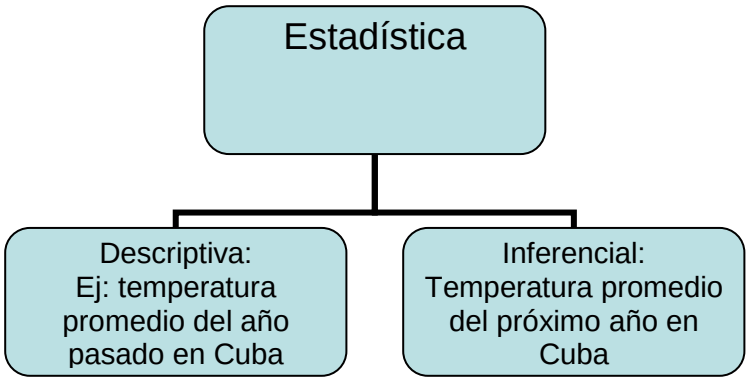
- Objetivo General desde la asignatura contribuir a la Formación Integral como Profesionales de la economía, acorde con las demandas de la sociedad cubana.
- Ayudarlos a que aprendan Estadística.

Presentación del Grupo. Utilizando una técnica participativa.

Escriban en una hoja los siguientes datos que servirán para introducir el tema y las variables:

Nombres y apellidos, Municipio, municipio de residencia, acceso a las TIC, en que emplea el Tiempo Libre, Número de hermanos, Nota final de Matemática Superior. Peso Aproximado.

La Estadística se divide en dos grandes ramas



Métodos Descriptivos: Describen el comportamiento de los datos estadísticos, se ocupan de: la recolección, organización, tabulación, presentación y reducción de la información.
Inferencia: Estudia y concluye sobre un fenómeno basado en el análisis e investigación de una parte del mismo.
Población: Colección de individuos o elementos que presentan el objeto de interés (seres vivos o inanimados).
Muestra: Cualquier subconjunto de la población.
Presentación de los alumnos, los datos anteriores me permiten introducir la clasificación de las variables en:
VARIABLE O CARACTERÍSTICA: Es el signo o detalle que interesa observar en la población. Variables cualitativas y cuantitativas la asignatura profundizara en estas ultimas.
VARIABLES DISCRETAS: Son valores, determinados, definidos y que generalmente representan a observaciones susceptibles de conteo y son valores enteros.
VARIABLES CONTINUAS: Las que generalmente representan a observaciones susceptibles de medición y por tanto pueden tomar cualquier valor en un cierto intervalo de los números reales.

1.2 .1. Distribución de frecuencia Univariada. Datos no agrupados. Propiedades de las Frecuencias. Datos Agrupados. Clases. Límites de clases. Tamaño del Intervalo de Clases. Marcas de Clases.

Datos Organizados son los que se ordenan en tablas llamadas de Distribuciones de frecuencias, que pueden estar agrupados y no agrupados.
Datos Organizados y no agrupados
EJEMPLO #1 Se tiene los datos recopilados acerca de la variable x: número de mantenimientos realizados en los 2 años a 20 mini industrias procesadoras de productos alimenticios.
0 1 2 2 1 3 2 1 4 2 **DATOS EN SU FORMA SIMPLE O PRIMARIA**
4 3 2 0 0 2 2 3 0 3 **(SIN ORGANIZAR**)
¿Qué tipo de variable es esta? **Discreta**.
Generalmente los datos que se organizan en tablas pero no se agrupan, son las variables discretas, sobre todo con fines teóricos, por su propia características, (Toma valores

determinados, definidos, susceptible de conteo y que por tanto generalmente son valores enteros).

TABLA DE FRECUENCIA O DISTRIBUCION DE FRECUENCIAS.

# de mantenimientos características		# de mini industrias Con esas			
Xi	tarjado	ni	fi	Ni	Fi
0	////	4	0.20	4	0.20
1	///	3	0.15	7	0.35
2	//// ///	7	0.35	14	0.70
3	////	4	0.20	18	0.90
4	//	2	0.10	20	1.00

Se definirá entonces:
ni = frecuencia absoluta: # de veces que se repite el í-esimo valor de la variable.
k = # de valores diferentes

$$\sum_{i=1}^{k} ni = n : \text{\# total elementos}$$

Interpretación: "ni" indica las veces que se repite el valor de la variable, así:
$n_1 = 4$ indica que hay 4 mini industrias del grupo que no recibieron mantenimientos.
$n_3 = 7$ indica que hay 7 mini industrias del grupo que recibieron 2 mantenimientos.

Este tipo de tabla se puede construir no solo con respecto a las frecuencias absolutas, sino también respecto a otros tipos de frecuencias

fi = frecuencia relativa : proporción de veces que se repite el í-esimo valor de la variable, donde fi = ni/n y donde $\sum fi = 1$.
N_3= 14 indica que 14 mini industrias han recibido a lo máximo 2 auditoría.

1.2.2. Datos Agrupados. Clases. Límites de clases. Tamaño del Intervalo de Clases. Marcas de Clases.

Ejercicio # 2 Los siguientes valores corresponden a las horas perdidas por roturas de 50 industrias del territorio nacional.

46	39	34	33	32	36	41	26	32	36
43	28	30	27	32	42	30	31	34	41
28	30	26	21	37	39	25	33	47	28
26	23	30	43	40	36	20	38	38	31
29	30	48	47	23	31	24	38	35	36

¿ Cómo se pueden agrupar estos datos en una tabla de frecuencias?
Primeramente se deben formar los intervalos de clases. No existe una regla fija de cuantos son los intervalos que se deben hacer; puede decirse que hay diferentes criterios, se tomará uno de ello, es que se deben considerar entre 5 y 20 intervalos.; ni muy pocos, que se pierda información, ni muchos que parezcan que no se han agrupados los datos.
Por tanto, ¿de qué depende? de la cantidad de observaciones, que se tienen, del tipo de información, de que se persigue con ella, etc.

El número de intervalos se representa por la letra "K"
¿Cuál debe ser la amplitud de estos intervalos?
Se considerará la misma amplitud para todos los intervalos y se representarán por "Ci" aunque, no necesariamente tiene que ser así.

a) Para llegar a ello se determina el **RECORRIDO DE LA VARIABLE o RANGO**.

R = Xmax. - Xmin. , que en este ejemplo será:

R = 48 - 20 = 28

b) Definir el número de intervalos. (K) recomendaciones para fijar el número de intervalos

Cantidad de Datos	cantidad de Intervalos (K)
20-50 datos	5-7
50-75	10
75-100	12
+100	15

Supóngase que se decide hacer K = 6 (número de intervalos), entonces la amplitud será:

Ci = R/K = 28 / 6 = 4.67 ≈ 5 redondeando, porque no tendría caso, hacer los intervalos de amplitud decimal, ya que lejos de facilitar el trabajo lo complicaría.
Y ya se está en condiciones de construir la tabla de frecuencia. Se debe comenzar por el menor valor, toda vez que es necesario que se incluyan todos los valores.

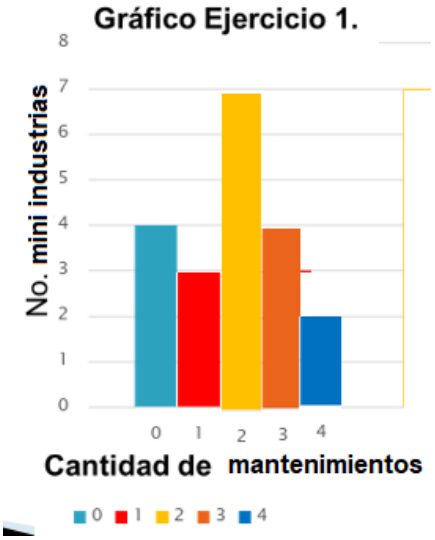
Se designará, como límite inferior del intervalo, por L_{i-1} y como límite superior Li.

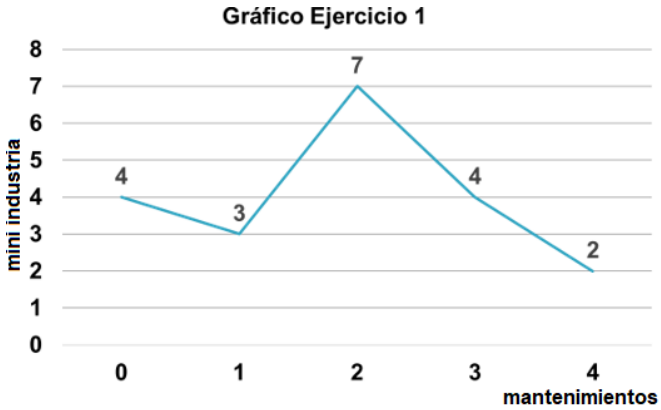
L_{i-1}	L_i	X_i	n_i	$X_i n_i$
20	25	22.5	6	135
25	30	27.5	13	357.5
30	35	32.5	11	357.5
35	40	37.5	11	412.5
40	45	42.5	5	212.5
45	50	47.5	4	190
			50	1665.0

La marca de clase (Xi) es el valor medio del intervalo

1.3 . Gráficos: Histograma y Polígono de frecuencias (Orientar).

Histograma





Polígono de frecuencias

1.4. Medidas Descriptivas de Posición, Tendencia Central más importantes: Moda, Mediana, Media. Sus propiedades.

MEDIDAS DE TENDENCIA CENTRAL O DE POSICION.

Existen 3 propiedades principales que describen un conjunto de datos numéricos, que son:
1.- Medidas de Posición (Ej. percentiles y de tendencia central (en este curso estudiaremos las de Tendencia Central.
2.- Dispersión
3.- Forma. (Ej. Curtosis. No estudiaran en este curso)

MEDIDAS DESCRIPTIVAS	Población (Parámetro)	Muestra (Estadígrafo)
Moda	**Mo**	**Mo**
Mediana	**Me**	**Me**
MEDIA	**μ**	$\overline{X}$

LA MODA

Se define como el valor mas frecuente, en un conjunto de datos. El valor modal es el de mayor frecuencia.
En datos organizados, donde exista un máximo relativo, esto es donde: $n_{i-1} < n_i > n_{i+1}$, existe un valor modal.

Se denota por Mo ó Md y puede no existir en una distribución, ó existir más de una.
Se utiliza en ocasiones como medida de tendencia central y se obtiene fácilmente a partir de un arreglo ordenado. A diferencia de la media. la moda no se afecta ante la ocurrencia de valores extremos.
Ejemplo en datos en su forma primaria: X = las calificaciones de 7 estudiantes del grupo.
4 3 4 5 4 5 3
Primeramente se ordenará de menor a mayor, esto es:
3 3 4 4 4 5 5 así el valor modal será el que más se repite Mo = 4

En el ejemplo #1 donde x: número de hermanos de 20 estudiantes del grupo 3210 (datos organizados, no agrupados, variable discreta)

Xi	ni	
0	4	Así el valor modal será el valor de "X" que le corresponde al máxi-
1	3	mo relativo.(buscado en las frecuencia absolutas).
2	7 *	Máximo relativo: nj-1 < nj > nj+1 así nj = 7. que es mayor que
3	4	nj-1 = 3 y mayor que nj+1 = 4
4	2	Así un máximo relativo será aquel valor que supera al que le antecede y al que le sigue en las frecuencias absolutas

Por lo tanto el valor modal será: Mo = 2

LA MEDIANA:

Se define como aquel valor que no supera ni es superado por más del 50% de las observaciones.

Se denota por Me

Para calcular la mediana, a partir de un conjunto de datos recolectados en su forma primaria, primero se deben ordenar de menor a mayor. A este conjunto se le denomina arreglo ordenado. Después se busca la posición del valor mediano en el arreglo ordenado. Se pudiera decir que hay 2 reglas, para determinar el valor mediano en los datos primarios. Y estas, están en dependencia del número de observaciones.

REGLA 1 Si el tamaño de la muestra es un número impar, la mediana está representada por el valor numérico correspondiente a la posición del centro de las observaciones ordenadas.
Ejemplo: si se tienen los valores siguientes:
5 8 3 4 6 2 3, primeramente lo que se hace es ordenarlo de menor a mayor: 2 3 3 4 5 6 8 y a continuación se obtiene el valor mediano el que será en este caso:
Me = 4
(por ser el valor del centro, que no supera ni es superado por el 50% de las observaciones
REGLA 2 Si el tamaño de la muestra es un número par, entonces la posición mediana, será la semi-suma de los dos valores centrales
Ejemplo: si se tienen los valores siguientes:

5 8 3 4 6 2, primeramente lo que se hace es ordenarlo de menor a mayor: 2 3 4 5 6 8 y el valor mediano será:

$$Me = 4 + 5/2 = 9/2 = 4.5$$

MEDIA ARITMETICA. También denominada ***media***, es el promedio ó medida de tendencia central, que se utiliza con mayor frecuencia.

Se define: como la suma de todos los valores de la variable entre el número de elementos. $\mu = \Sigma Xi / n$

Se representa: En la muestra por $\overline{X}$, y en la población por μ ó M(x).

Forma de cálculo: Datos primarios --- $\overline{X} = \sum_{i=1}^{n} Xi / n$

Estudiar sus propiedades: (Libro de texto: pág. 37-39)

A continuación se verá el ejemplo que se realizó para datos organizados, pero sin agrupar, para la variable discreta: X = número mantenimientos de n=20 mini industrias del grupo.

Xi	ni	Xini	fi	Xifi
0	4	0	0.20	0
1	3	3	0.15	0.15
2	7	14	0.35	0.70
3	4	12	0.20	0.60
4	2	8	0.10	0.40
	20	37	1.00	1.85

$$\overline{X} = \sum_{i=1}^{n} Xini/n = 37/20 = 1.85$$

$$\overline{X} = \sum_{i=1}^{n} Xifi = 1.85$$

1.5. Medidas de Dispersión:

Es necesario destacar que las medidas de tendencia central, nos describen solo una parte del comportamiento de los datos. Para tener una descripción más detallada es necesario que estén acompañadas de medidas de dispersión, porque a través de ellas se puede determinar si la medida de tendencia central es representativa de la distribución.

Ejemplo:

Deseamos calcular salario promedio de 3 empleados (A, B y C) durante siete días

.Muestra A: 55 55 55 54 56 55 55

Muestra B: 47 51 53 55 57 59 63

Muestra C: 39 47 53 55 57 63 71

En todos los casos la Media es 55 cup, sin embargo, ese estadístico es más representativo de la muestra A que de los restante porque en, A está más cerca de todos sus valores

Dentro de los estadísticos de dispersión están **Rango, Varianza y Desviación Típica y coeficiente de variación**.

RANGO

Ya se estudio $R = Xmax - X\ min$

LA VARIANZA

La varianza de un conjunto de datos x1, x2, x3, ... , xn se define como la media aritmética del cuadrado de las desviaciones de la variable con respecto a su media.

Se denota por S^2 en la muestra y por σ^2 en la población o también por V(x).

Expresión matemática:

$$S^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2 n_i / n$$

Veamos el cálculo en datos organizados en el ejemplo #1 donde x: es el número de mantenimientos realizadas 20 mini industrias del territorio.

Xi	n_i	$(Xi- \bar{X})$	$(Xi- \bar{X})^2$	$(Xi- \bar{X})^2 n_i$
0	4	(0 -1.85)	3,4225	13,6900
1	3	(-0.85)	0,7225	2,1675
2	7	(0.15)	0,0225	0,1575
3	4	(1.15)	1,3225	5,2900
4	2	(2.15)	4,6225	9,2450
				∑=30,5500

S^2= 30.55/20 = 1.53 mantenimientos^2

Este resultado no se puede interpretar por estar las cantidades elevadas al cuadrado.

Otra forma de calcular la varianza es mediante la formula

$$S^2 = \sum_{i=1}^{n} (Xi^2 ni)/n - \bar{X}^2$$

Veamos el cálculo en datos organizados en el ejemplo #1 donde x: es el número de mantenimientos realizadas 20 mini industrias del territorio.

Xi	ni	Xini	Xi^2ni
0	4	0	0
1	3	3	3
2	7	14	28
3	4	12	36
4	2	8	32
//	20	37	99

donde $\bar{X}$ = 1.85 por tanto es más conveniente usar la formula:

$$S^2 = [(\sum_{i=1}^{n} Xi^2 ni) / n] - (\bar{X})^2$$

que es más cómoda, menos complicada de trabajar

$S^2 = 99/20 - 1.85^2 = 4.95 - 3.42 = 1.53$

$S^2 = 1.53$ Este resultado no se puede interpretar por estar las cantidades elevadas al cuadrado.

$$S^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2 n_i / n$$

La expresión de la varianza cambia en el denominador para la población y la muestra

$$\sigma^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2 ni}{n} \qquad S^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2 ni}{n-1}$$

Estudiar las Propiedades de la Varianza. (LT. p.45) y Guía en Formato Digital Carpeta Contenidos/ Tema 1

La varianza como se definió es el promedio de las desviaciones de la variable con respecto a su media elevadas al cuadrado

Mientras mayor sea la varianza más dispersos estarán los valores alrededor de la media, y mientras más pequeña, menos dispersión habrá, pero estas son cantidades elevadas al cuadrado y no se puede interpretar como tal y tampoco da una idea clara.

Entonces buscaremos otro Estadístico.

LA DESVIACION TIPICA.

La desviación típica se define como la raíz cuadrada positiva de la varianza

Se denota por S en la muestra y por σ en la población.

$S = + \sqrt{S^2}$ La desviación típica es una magnitud no negativa, pero no cumple las restantes propiedades de la Varianza, pues la extracción de la raíz no lo permite.

Así en los ejemplos vistos en datos organizados será:
Ejemplo #1 x: número de mantenimientos realizadas a 20 mini industrias del territorio.

$S^2 = 1.53$ lo que implica que $S = \sqrt{1.53} = 1.23$
Interpretación: que en promedio los datos se desvían de la media en 1.23 unidades.
Es decir la desviación típica está en las mismas unidades en que está expresada la variable. Esta dado en valores absolutos.
En ocasiones resulta necesario contar con un estadístico que refleje la dispersión sin depender de la magnitud de las observaciones, esto es que sea un valor relativo.
Esta necesidad surge generalmente cuando se comparan las dispersiones entre varios conjuntos.
Ejemplo
Analicemos la variabilidad en el costo de producción de un artículo dado, en dos fábricas diferentes, ¿Cuál considera que tiene mayor dispersión en sus producciones?

Fábrica A: Media (μ)= 5.85 pesos y Desviación Típica(σ)= 1.25.
Fábrica B: Media (μ)= 5.15 pesos y Desviación Típica(σ)= 1.20.

Podría parecer que la Fábrica A tiene mayor variabilidad que la B hagamos el análisis a partir de otro Estadístico el Coeficiente de Variación y que se expresa en valores relativos.

COEFICIENTE DE VARIACION.

Se define como el cociente entre la desviación típica y la media multiplicada por 100, para dar el resultado en porciento.

Se denota por C.V. $C.V = S/X * 100$

Este coeficiente es una medida de dispersión igual que la desviación típica (S), únicamente que éste es una medida relativa, mientras que S es una medida absoluta.
Así el resultado que se alcanzaría en los dos ejemplos que se han venido realizando para datos organizados no agrupados y agrupados será de:
Ejemplo #1 donde x: número de mantenimientos realizadas a 20 mini industrias del territorio.

C.V. = 1.23 /1.85 * 100 = 0.66 * 100 = 66%

Los datos están dispersos como promedio con respecto a la media en un 66%
Es bastante alta la dispersión, sin embargo si se observa S = 1.23 se hubiese podido pensar que era baja.

A la luz de este Estadístico les proponga ahora que determinen en cúal de las dos fábricas es mayor la dispersión de los costos con relación a su media.

1.6 Medidas descriptivas de formas: Asimetría y Curtosis

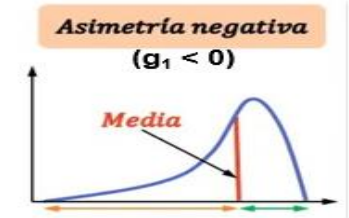
Asimetría

El *Coeficiente de Asimetría* (g_1) permite identificar si los datos se distribuyen de forma uniforme alrededor del punto central (Media aritmética).

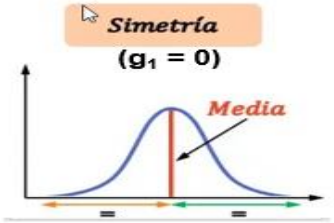
La asimetría presenta tres estados diferentes:

- asimetría negativa.
- simétrica.
- asimetría positiva.

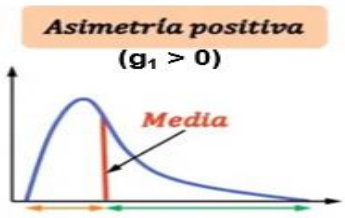
Desde luego entre mayor sea el número (Positivo o Negativo), mayor será la distancia que separa la aglomeración de los valores con respecto a la media



Cuando la mayoría de los datos se distribuyen por debajo del valor de la media. **($g_1 < 0$)**



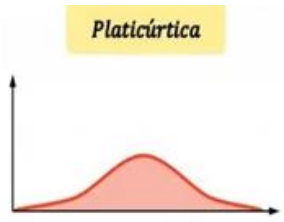
Cuando se distribuyen aproximadamente la misma cantidad de valores en ambos lados de la media. **($g_1 = 0$)**



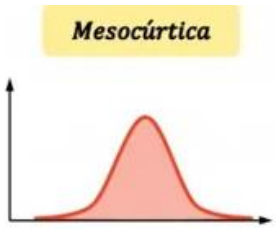
Cuando la mayoría de los datos se distribuyen por encima del valor de la media. **($g_1 > 0$)**

Curtosis

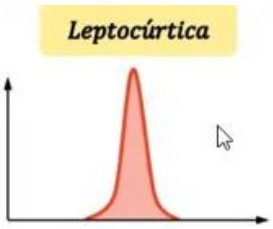
El Coeficiente de Curtosis (g_2), determina el grado de concentración que presentan los valores en la región central de la distribución



Platicúrtica: ($g_2 < 0$)



Mesocúrtica: ($g_2 = 0$) suelen aceptar los valores cercanos (± 0.5 aprox.).



Leptocúrtica: ($g_2 > 0$)

1.7 Medidas Descriptivas datos cualitativos: Razón, Proporción y Porcentaje.

Ej. Se realizó la inspección a 120 mini industrias para comprobar si cumplen sus planes de negocios.

a: 100 mini industrias cumplen,

b: 20 mini industrias no cumplen**.**

Razón

$$R = a / b$$

Razón de mini industrias que cumplen

$$R = 100 / 20 \ldots\ldots\ldots 5/1 = 5$$

Hay 5 que cumplen su plan de negocio por cada 1 que no lo cumple

Proporción

$$P = a/n$$

a: característica que se quiere medir (cumplen el plan); n: número total de mini industrias

$$P = 100/120 = 5/6 = 0.833$$

Hay 5 mini industrias que cumplen el plan de cada 6 mini industrias inspeccionada

Porcentaje

$$\% = (a / n) * 100$$

$$\% = P * 100$$

Porcentaje de mini industrias que cumplen el plan

$$(100/120)*100= 83,3\%$$

El 83,3% de las mini industrias inspeccionadas cumplen el plan. Aproximadamente 83 de cada 100 mini industrias inspeccionadas cumplen sus planes de negocios

¿Qué porciento no cumplen el plan?

Conclusiones:

Resumiendo lo tratado, con énfasis en:

- Clasificación de las variables.

- Se presenta un cuadro con los parámetros y estadísticos fundamentales de variables cuantitativas

PARAMETRO O ESTADÍSTICO	PARAMETRO PARA LA POBLACIÓN	ESTADÍSTICO PARA LA MUESTRA
MODA	Mo	Mo
MEDIA	Me	Me
MEDIA	μ	$\overline{X}$
VARIANZA	$\sigma 2$	S2
DESVIACIÓN TÍPICA	σ	S

Estudio Independiente No 1.

1. Construya la Tabla de Distribución de Frecuencias absolutas y relativas por intervalo del Ejercicio 2.
2. Construya el Histograma de frecuencias representando en el eje de las abscisas la Marca de clase.

Se enviará las preguntas de la Clase Práctica 1 para que los estudiantes se auto preparen.